HIGH THROUGHPUT SEQUENCING AND ANALYZING OF THE CDNAS OF HYPOTHETICAL GENES OF ARABIDOPSIS ON CHROMOSOME 2.

Yong-Li Xiao, Shannon R Smith, Nadeeza Ishmael, Julia Redman, Nihkil Kumar, Erin Monaghan, Mulu Ayele, Brian Haas, Christopher D Town
The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850
yxiao@tigr.org

In fully sequenced Arabidopsis genome, about 20% of genes are called "hypothetical gene". This particular group of genes has structures that are predicted solely by computer algorithms with no support from either nucleic acid or protein homologs from other species or expressed sequence matches from Arabidopsis. A high throughput method was developed to clone and analyze cDNAs of these genes. Primers from 797 hypothetical genes on chromosome 2 were designed and clones from 5' or 3' RACEs of 506 genes were obtained and sent to sequence. The generated 11327 sequences match 399 genes from our target list. The cDNA sequences were obtained by sequencing and assembling 5' and 3' RACE products, which displays that the structures of 151 hypothetical genes are different with their predicted gene structures. In total, 87 hypothetical genes have alternatively spliced transcripts and 110 genes display more than one polyadenylation site. In addition, transcripts from opposite strands of their predictions, transcripts from both strands at same regions and dicistronic transcripts were found. The cDNA populations used in this study are from cold, heat, 2,4-D, H_2O_2, UV, IAA, salt-treated and pathogen (*Xanthomonas campestris* pv. campestris and *Pseudomonas syringae*)-infected plants, callus, roots and young seedlings. The promoter regions of 5 hypothetical genes were cloned into GFP and GUS reporter constructs and their expression patterns show tissue and development stage specific.